

RESEARCH

Open Access



Construction and validation of a cuproptosis-related diagnostic gene signature for atrial fibrillation based on ensemble learning

Yixin Wang¹, Qiaozhu Wang¹, Peng Liu¹, Lingyan Jin¹, Xinghua Qin^{2*} and Qiangsun Zheng^{1*}

Abstract

Background Atrial fibrillation (AF) is the most common type of cardiac arrhythmia. Nonetheless, the accurate diagnosis of this condition continues to pose a challenge when relying on conventional diagnostic techniques. Cell death is a key factor in the pathogenesis of AF. Existing investigations suggest that cuproptosis may also contribute to AF. This investigation aimed to identify a novel diagnostic gene signature associated with cuproptosis for AF using ensemble learning methods and discover the connection between AF and cuproptosis.

Results Two genes connected to cuproptosis, including solute carrier family 31 member 1 (SLC31A1) and lipoic acid synthetase (LIAS), were selected by integration of random forests and eXtreme Gradient Boosting algorithms. Subsequently, a diagnostic model was constructed that includes the two genes for AF using the Light Gradient Boosting Machine (LightGBM) algorithm with good performance (the area under the curve value > 0.75). The microRNA-transcription factor-messenger RNA network revealed that homeobox A9 (HOXA9) and Tet methylcytosine dioxygenase 1 (TET1) could target SLC31A1 and LIAS in AF. Functional enrichment analysis indicated that cuproptosis might be connected to immunocyte activities. Immunocyte infiltration analysis using the CIBERSORT algorithm suggested a greater level of neutrophils in the AF group. According to the outcomes of Spearman's rank correlation analysis, there was a negative relation between SLC31A1 and resting dendritic cells and eosinophils. The study found a positive relationship between LIAS and eosinophils along with resting memory CD4⁺ T cells. Conversely, a negative correlation was detected between LIAS and CD8⁺ T cells and regulatory T cells.

Conclusions This study successfully constructed a cuproptosis-related diagnostic model for AF based on the LightGBM algorithm and validated its diagnostic efficacy. Cuproptosis may be regulated by HOXA9 and TET1 in AF. Cuproptosis might interact with infiltrating immunocytes in AF.

Keywords Atrial fibrillation, Cuproptosis, Diagnostic signature, Bioinformatics, Ensemble learning

*Correspondence:

Xinghua Qin
xinghuaqin@nwpu.edu.cn
Qiangsun Zheng
zhengqiangsun@126.com

¹Department of Cardiology, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China

²Xi'an Key Laboratory of Special Medicine and Health Engineering, School of Life Sciences, Northwestern Polytechnical University, Xi'an, China



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Atrial fibrillation (AF) is a common cardiac arrhythmia in healthcare facilities, with a global prevalence exceeding 43 million individuals [1]. AF is a substantial risk factor for ischemic stroke, as it increases the probability of stroke by five times and is responsible for approximately one-third of all strokes [2–4]. Furthermore, strokes among subjects with AF are connected with elevated mortality compared to strokes in individuals without AF [5]. It is known that the administration of oral anti-coagulation (OAC) could significantly mitigate the risk of AF-related stroke [6, 7]. However, since AF can escape traditional monitoring techniques due to its often asymptomatic and paroxysmal nature [8], leading to delayed onset of OAC, ischemic stroke is often the initial sign of AF [9]. Therefore, novel diagnostic approaches supplementing the current methods for the timely detection of AF are urgently required.

The pathophysiological pathways for the beginning and perpetuation of AF are extremely complex. There is a growing body of evidence suggesting that genetic factors are a significant contributor to the development of AF. Genome-wide association studies have identified approximately 140 genetic loci that are associated with AF [10]. Recently, the rapid advancement in microarray technology has enabled the identification of gene biomarkers associated with AF [11, 12], enabling the development of new diagnostic models based on genes for diagnosing AF. Additionally, electrical and structural remodeling, the predominant mechanism underlying AF, has been associated with various types of cell death, such as ferroptosis, necroptosis, apoptosis, and autophagy [13–16]. Recently, Tsvetkov et al. introduced a new form of cellular death known as cuproptosis, which is triggered by excessive accumulation of copper (Cu) [17]. Grandis et al. reported that Wilson's disease (WD), a disease resulting from abnormal Cu metabolism, induced a higher risk of AF [18], which may be a consequence of myocardial Cu deposition [19]. Significantly, Cu is involved in immunity [20].

The immune response and inflammation are two crucial mechanisms in AF pathogenesis. Numerous inflammatory biomarkers, including interleukins, C-reactive protein, and tumor necrosis factor- α , have been associated with AF [21]. Relevant studies revealed a vital role of the immunocyte infiltration of atrium in the pathogenesis of AF [22]. In summary, it is reasonable to consider that cuproptosis is tightly connected to the pathogenesis of AF. Therefore, the establishment of a gene signature linked to cuproptosis may provide the foundation to further investigate the association between AF and cuproptosis, which can shed new light on the diagnosis and management of individuals with AF.

Machine learning (ML), involving procedures that learn to make decisions from data, demonstrated success and scalability in the diagnosis and prognosis of AF [23]. Ensemble learning (EL) is a subfield of ML. The utilization of EL algorithms in computational biology has become more prevalent as a result of their distinct benefits in managing limited sample sizes, intricate data constructions, and high-dimensional data [24]. To our knowledge, the application of EL algorithms such as Random Forests (RF) [25], eXtreme Gradient Boosting (XGBoost) [26], and Light Gradient Boosting Machine (LightGBM) [27] has not yet been reported for the diagnostic gene signature of AF. RF algorithm allows predictors to be ranked according to their importance in a regression or classification problem [28]. XGBoost and LightGBM algorithms are both based on gradient-boosting tree-based methods. XGBoost could analyze feature importance internally throughout the learning process and provide scores for all features. The LightGBM algorithm exhibits superior performance compared to XGBoost, with notable enhancements in performance, training speed, and accuracy [29].

This study aims to discover the association between cuproptosis-associated genes and AF, investigate the diagnostic importance of cuproptosis-associated gene signature based on the LightGBM algorithm, study the correlations between cuproptosis and immunocyte infiltration, and construct the microRNA (miRNA)-transcription factor (TF)-messenger RNA (mRNA) regulatory network of the genes. The analysis process of this investigation is illustrated below (Fig. 1).

Materials and methods

Acquisition and preprocessing of datasets

This study employed “atrial fibrillation” as the designated keyword, specified the study organism as “Homo Sapiens”, and identified the study type as “Expression profiling by array”. Subsequently, it conducted a thorough exploration of the atrial fibrillation-associated datasets within the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>). Finally, the matrix of expression for five unique datasets, namely GSE79768, GSE31821, GSE41177, GSE14975, and GSE115574 (Table 1), was obtained from the GEO repository. Samples of atrial tissue from sinus rhythm (SR) controls and patients with AF were selected for analysis. The Affymetrix Human Genome U133 Plus 2.0 Array (GPL570) was utilized to annotate all five datasets. The conversion of probes in each dataset was performed utilizing annotation files with ActivePerl (5.18.4) to obtain corresponding gene symbols. GSE14975 was first transformed into log₂ transformed first to ensure consistency with the other four datasets, which had been preprocessed with log₂ transformation before. In order to combine these

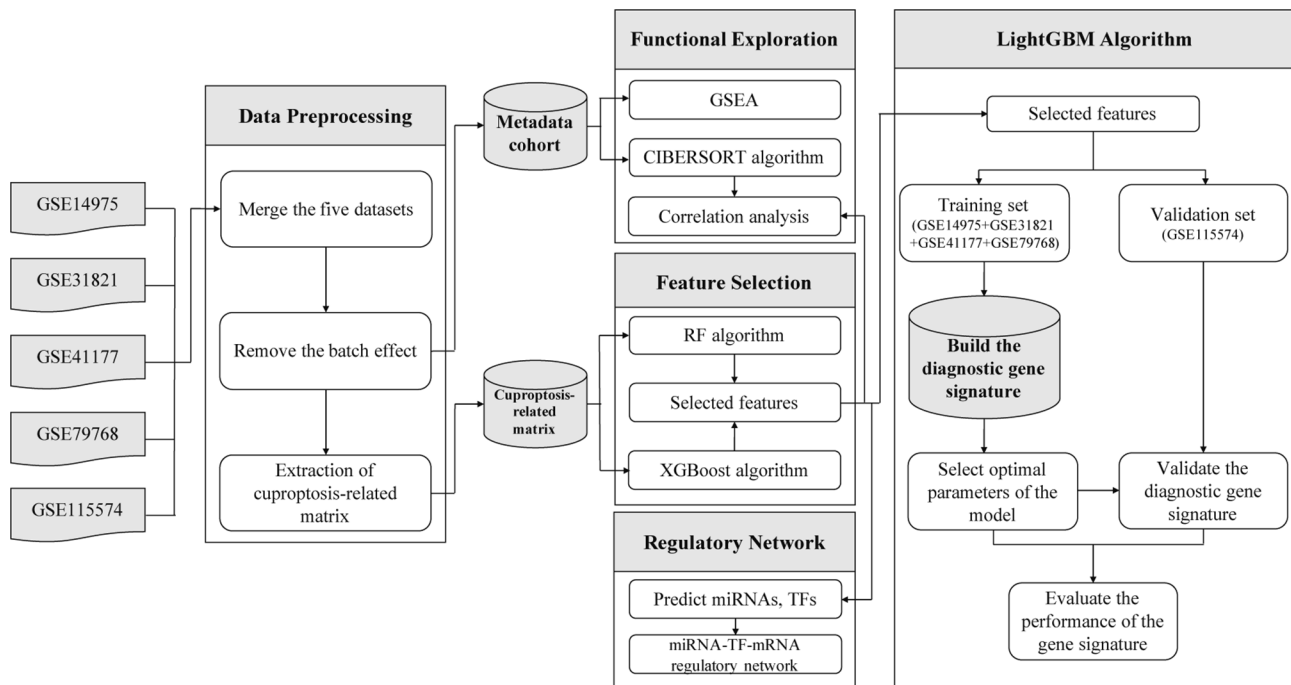


Fig. 1 Flowchart of this study

Table 1 Details of the five datasets

Series accession	Experiment type	Sample size (selected/total)	Sample size (AF/SR)	Contributors	Country	Last update date
GSE14975	Expression profiling by array	10/10	5/5	Adam O, et al.	Germany	2019/3/25
GSE31821	Expression profiling by array	5/6	3/2	Morel E, et al.	France	2019/3/25
GSE41177	Expression profiling by array	38/38	32/6	Yeh Y, et al.	China (Taiwan)	2019/3/25
GSE79768	Expression profiling by array	26/26	14/12	Tsai FC, et al.	China (Taiwan)	2019/3/25
GSE115574	Expression profiling by array	59/59	28/31	Deniz GC, et al.	Turkey	2021/12/13

five datasets as a metadata cohort, the batch effect should be removed. Batch normalization was performed for the merged expression data of five datasets in R (4.1.0) with the “sva” package(3.40.0) [30], and the ComBat method was used to normalize the expression values from different datasets [31]. Table S1 presents that 13 cuproptosis-related genes were collected from prior research [17]. Cuproptosis-related gene matrix was extracted from the metadata cohort based on the cuproptosis-related genes using R (4.1.0).

Cuproptosis-related gene selection utilizing RF and XGBoost algorithms

To identify cuproptosis-related diagnostic variables tightly related to AF, RE, and XGBoost, algorithms were implemented in the cuproptosis-related matrix. RF algorithm was conducted to compute the importance score with 1000 classification trees constructed initially using the “randomForest” package (4.6–14) [28]. Subsequently, the optimal number of trees to grow (ntree) was determined according to the minimum error rate. The

features with Gini importance ranked as the top 3 were considered. XGBoost algorithm was implemented with parameters set as “the learning rate (eta)=0.3, maximum depth of a tree (max_depth)=6, max number of boosting iterations (nrounds)=10” through the “xgboost” package (1.5.0.2, <https://github.com/dmlc/xgboost>). The features with relative Gain-importance ranking top 3 were selected. Finally, the overlapping cuproptosis-related genes from the two algorithms were selected to establish the diagnostic gene signature, and the Venn diagram was produced utilizing VENN 2.1 (<https://bioinfogp.cnb.csic.es/tools/venny/>).

The diagnostic gene signature construction and validation

The present study artificially split the cuproptosis-related matrix into two sets: the training set containing GSE14975, GSE31821, GSE41177, and GSE79768, and the validation set containing GSE115574. The cuproptosis-related genes selected by RF and XGBoost algorithms were submitted to the LightGBM algorithm to build the diagnostic gene signature in training set

utilizing the “lightgbm” package (3.3.2, <https://github.com/Microsoft/LightGBM>). The optimal parameters of a LightGBM-based model, including minimal sum Hessian in one leaf (`min_sum_hessian_in_leaf`), L1 regularization (`lambda_l1`), L2 regularization (`lambda_l2`), and the ratio of structures randomly chosen on every iteration (`feature_fraction`), were determined based on the minimum square loss during the process of training. Finally, the diagnostic gene signature was built with the optimal parameters and other parameters set as “`eta=0.1`, `rounds=100`” and validated in the validation set.

Evaluation of the LightGBM-based Diagnostic Gene signature

The efficiency of the diagnostic gene signature according to LightGBM was assessed utilizing the area under the curve (AUC) of receiver operator characteristic (ROC) and precision-recall (PR) curves. Specifically, the ROC-AUC and PR-AUC were utilized for this purpose. ROC curves were generated by the “pROC” package (1.18.0) [32], while PR curves were developed through the “ggplot2” package [33]. In general, an AUC value > 0.75 was used as a threshold for good discriminating capacity.

Functional Enrichment Analysis through Gene Set Enrichment Analysis (GSEA)

In the metadata cohort, GSEA [34] was implemented in the metadata cohort through the “clusterProfiler” package (4.0.5) [35] to discover the functional divergence between AF and SR. Gene ontology (GO)-biological process (BP), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Hallmark enrichment analysis were performed, respectively. The reference gene sets “`h.all.v7.5.1.symbols.gmt`”, “`c2.cp.kegg.v7.5.1.symbols.gmt`”, and “`c5.go.v7.5.1.symbols.gmt`” were obtained from Molecular Signatures Database (MSigDB, <https://www.gsea-msigdb.org/gsea/msigdb/index.jsp>). The significance threshold was established as an adjusted P-value ($adj.p$) of less than 0.05 and a false discovery rate (FDR) of less than 0.25.

Interactions between cuproptosis-related genes and immunocyte infiltration

The CIBERSORT algorithm [36] was performed to detect the relative proportions of 22 forms of infiltrating immunocytes (LM22) in patients with AF. The CIBERSORT algorithm was run using the LM22 gene set at 1000 permutations. The $p < 0.05$ served as the criteria for the inclusion of samples. The Wilcoxon assessment was adopted to compare the variations in the proportions of immunocytes among AF patients and SR controls. A statistical significance level of $p < 0.05$ was considered acceptable. Correlation analysis between the cuproptosis-associated genes selected by EL algorithms and Spearman’s rank

correlation analysis was utilized to conduct an immune response. Given that the $|\text{correlation coefficient } (R)| < 0.2$ indicates no correlation [37], it is necessary to set the criterion for the significance of correlation analysis as $|R| \geq 0.2$ and $p < 0.05$.

Construction of MicroRNA (miRNA)-Transcription factor (TF)-messenger RNA(mRNA) network

The genes associated with Cuproptosis, as identified by two EL algorithms, were subjected to analysis utilizing version 1.14.0 of the “multiMiR” package [38]. This analysis aimed to identify miRNAs in verified miRNA-target databases (miRecords, miRTarBase, and TarBase) as well as anticipated miRNA-target databases (DIANA-microT, EIMMo, MicroCosm, miRanda, miRDB, PicTar, PITA, and TargetScan), respectively. To improve the accuracy of prediction, only miRNAs fished by at least three predicted or two verified databases were retained. Meanwhile, TFs targeting cuproptosis-related genes were predicted using the “TF perturbations followed by expression table” module in Enrichr (<https://maayanlab.cloud/Enrichr/>) [39]. The $adj.p < 0.05$ served as a criterion for the inclusion of TFs, and the miRNA-TF-mRNA network was visualized and analyzed by the Cytoscape program (3.8.2) after identifying the modulatory relationships of miRNA-TF-mRNA [40].

Statistical analysis

The statistical analyses were performed utilizing R (version 4.1.0). A significant difference is typically denoted by $p < 0.05$.

Results

Batch normalization of data

In the analysis of five datasets, principal component analysis (PCA) was applied to investigate the sample clustering patterns before and after batch normalization. Before the batch normalization, the specimens were collected in batches depending on the top two principal components (PCs, Fig. 2A). Conversely, the scatter plot of normalized data suggested that the batch effect was successfully removed (Fig. 2B).

Progression of the cuproptosis-associated diagnostic gene signature

The cuproptosis-related matrix contained 12 cuproptosis-related genes since glycine cleavage system protein H (GCSH) did not exist in the metadata cohort. RF algorithm identified three features, including lipoic acid synthetase (LIAS), ATPase copper transporting Alpha (ATP7A), and solute carrier family 31 member 1 (SLC31A1) (Fig. 3A–B). Meanwhile, XGBoost selected three cuproptosis-related genes, including SLC31A1, LIAS, and dihydrolipoamide s-succinyltransferase

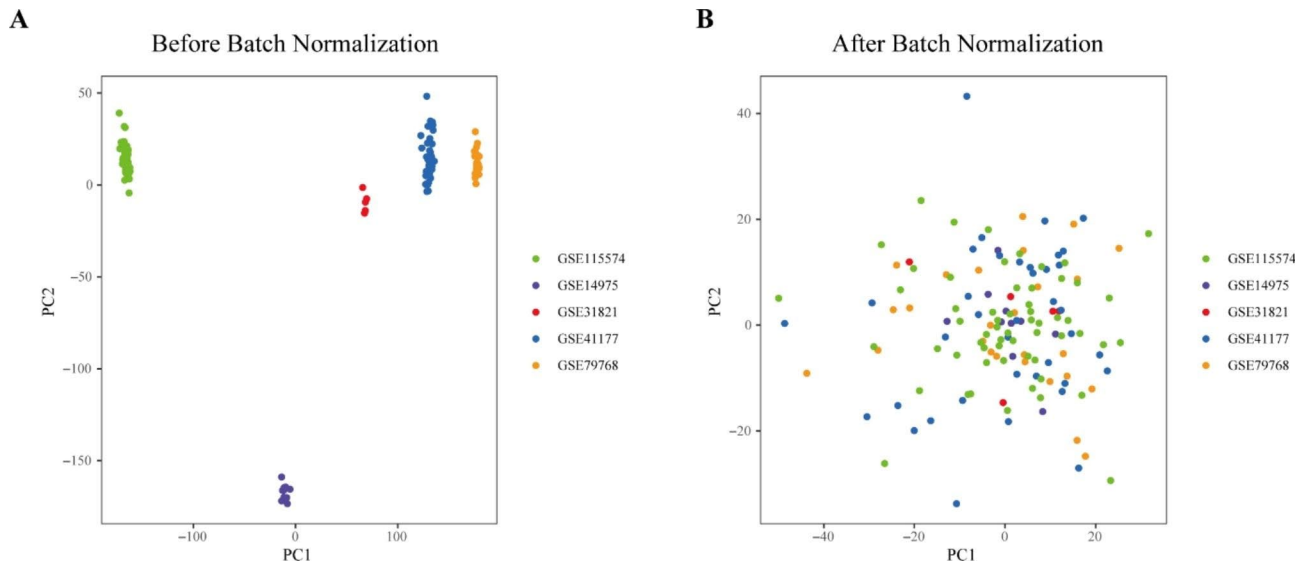


Fig. 2 Scatter-plot of PCA. **(A)** Before batch normalization. **(B)** After batch normalization

(DLST, Fig. 3C). The two overlapping features (SLC31A1, LIAS), which were tightly related to AF, were ultimately selected to build the diagnostic gene signature (Fig. 3D). During the model training procedure, the optimal parameters of LightGBM-based model were finally determined as “lambda_l1=0, lambda_l2=1, min_sum_hessian_in_leaf=0, feature_fraction=0.8”. The diagnostic gene signature with optimal parameters was validated in the validation set.

Diagnostic efficacy of the cuproptosis-related diagnostic gene signature

AUC-ROCs only compare the true- and false-positive rates, which means that AUC-ROCs only depict the capability of signature to discriminate between AF and SR. However, the signature is expected to have better performance in identifying AF but not SR in clinical scenarios. Consequently, PR curves comparing true and predicted positives were employed to assess the signature performance. The value of ROC-AUCs and PR-AUCs in both training and validation sets was higher than 0.75 (Fig. 4A–B), indicating that the signature had a good utility for discriminating between AF and SR, and performed a good separation that specifically mapped to AF.

Functional enrichment analysis

To comprehensively understand the variations in gene roles and mechanisms between groups characterized by the cuproptosis-related diagnostic gene signature, GSEA was conducted. Since the cuproptosis-related diagnostic gene signature displayed a good separation specifically mapped to AF, the functional annotations enriched in the AF group were valued. In GO-BP enrichment analysis, BP terms were significantly enriched in the immune

response, including activation of the immune response, immune response-regulating signaling mechanism, immune response-regulating cell surface receptor signaling pathway, positive regulation of immune response, and leukocyte migration (Fig. 5A). KEGG enrichment analysis revealed that remarkably enriched pathways in AF were mainly immunocyte-related, such as Fc gamma receptor (Fc gamma R)-mediated phagocytosis, chemokine signaling mechanism, intestinal immune network for immunoglobulin A (IgA) production, Leishmania infection and lysosome (Fig. 5B). Meanwhile, Hallmark terms significantly enriched in AF were allograft rejection and complement, which were tightly related to immunity (Fig. 5C).

In summary, the findings demonstrate that the cuproptosis-related diagnostic signature may be tightly related to the biological activities of immunocytes, which has an indispensable function in AF pathogenesis.

Immunocyte infiltration and correlation analysis

Depending on the functional enrichment analysis outcomes, the CIBERSORT procedure was employed to quantify the composition of 22 forms of immunocytes between the AF and SR groups categorized by the cuproptosis-related diagnostic gene signature. The outcomes exhibited that neutrophil infiltration was significantly elevated in AF patients (Fig. 6A–B). Moreover, correlation analysis was conducted between the two cuproptosis-related genes belonging to the diagnostic signature and infiltrating immunocytes. SLC31A1 was negatively associated with resting dendritic cells ($R = -0.27$, $p = 0.015$) and eosinophils ($R = -0.28$, $p = 0.012$, Fig. 6C and 7A–B). LIAS was positively correlated with eosinophils ($R = 0.24$, $p = 0.028$) and resting memory CD4⁺ T cells ($R = 0.37$,

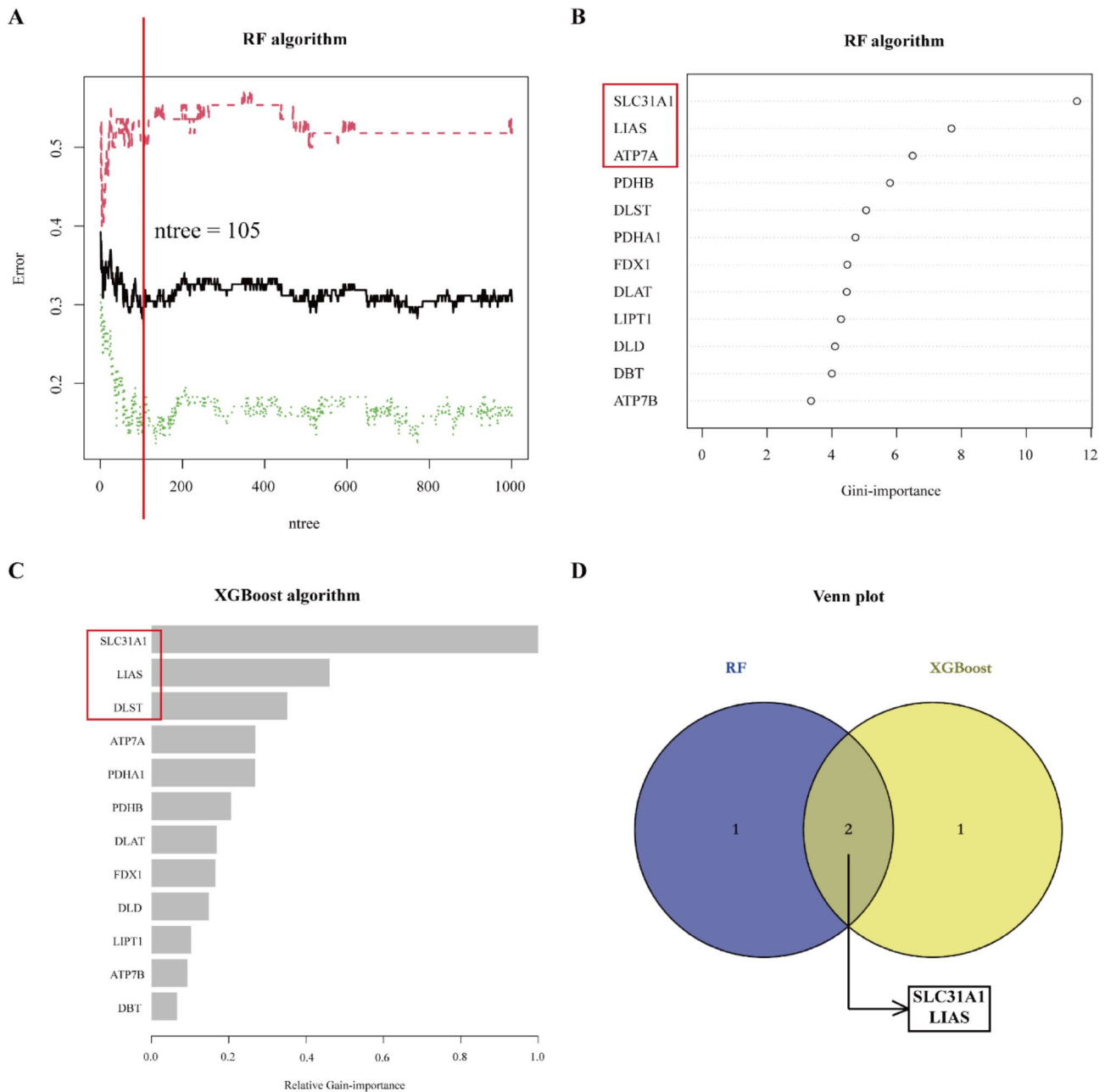


Fig. 3 Feature selection with RF and XGBoost algorithms. **(A)** Relationship between the error rate and the number of classification trees. The error rate is minimum when ntree=105. **(B)** Gini-importance of the 12 cuproptosis-related genes. **(C)** Relative Gain-importance of the 12 cuproptosis-related genes. **(D)** Venn plot demonstrating two features shared by RF and XGBoost algorithms

$p=8 \times 10^{-4}$), but negatively related to CD8⁺ T cells ($R = -0.3$, $p=0.0075$) and regulatory T cells (Tregs, $R = -0.27$, $p=0.015$, Figs. 6D and 7 C–F).

Regulatory Network of cuproptosis-related genes
 The miRNA-TF-mRNA network containing 2 mRNAs, 22 miRNAs, and 22 TFs was constructed (Fig. 8). In the network, SLC31A1 mRNA and LIAS mRNA were targeted by homeobox A9 (HOXA9) and Tet methylcytosine dioxygenase 1 (TET1).

Discussion
 Multiple mechanisms, including genetic factors, various types of cell death, immunocyte infiltration, and inflammation, are connected to the incidence and development of AF. Up to now, the diagnosis of AF has still been a challenge because it is often paroxysmal and asymptomatic in clinics. Gaining more insight into the underlying mechanism of AF would enable novel methods to diagnose AF. Tsvetkov et al. reported that the binding of Cu to lipoylated ingredients in the tricarboxylic acid cycle

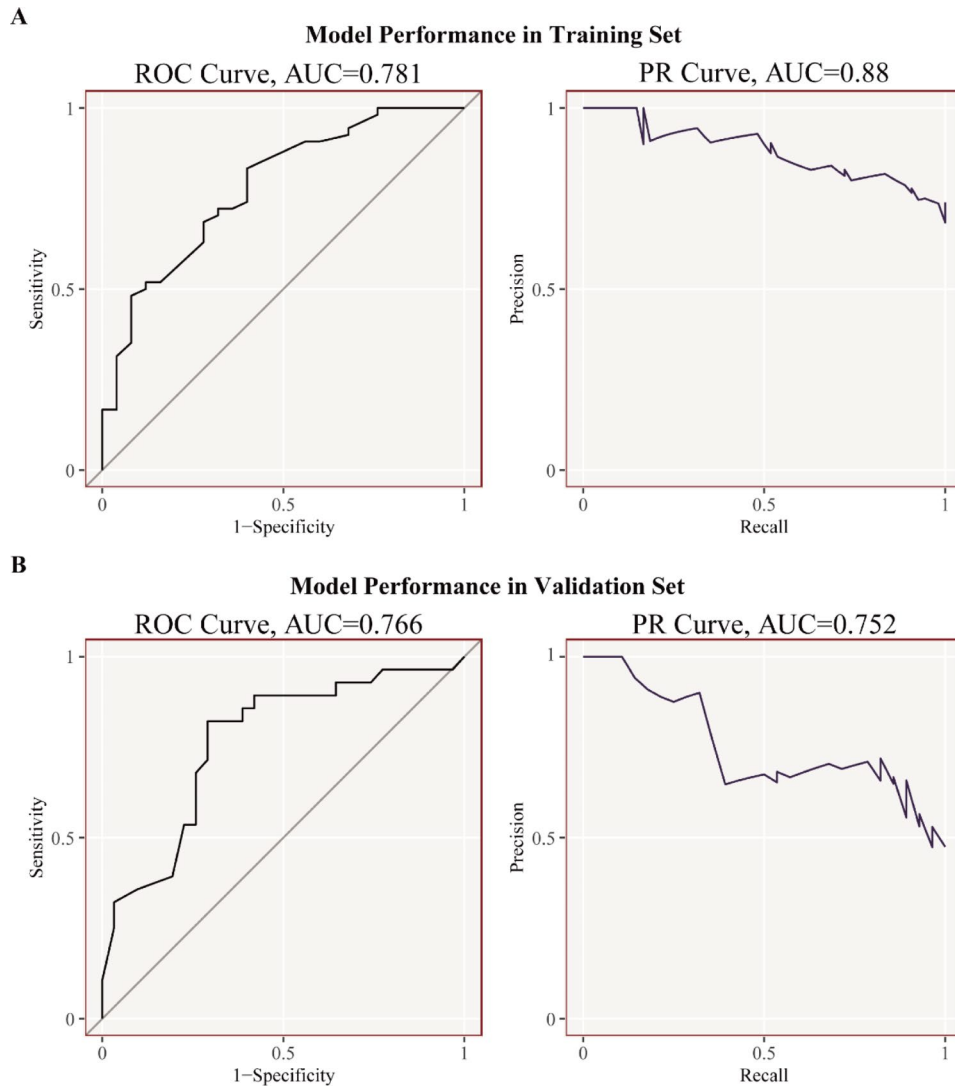


Fig. 4 PR and ROC curves of the diagnostic gene signature. **(A)** Training set. **(B)** Validation set

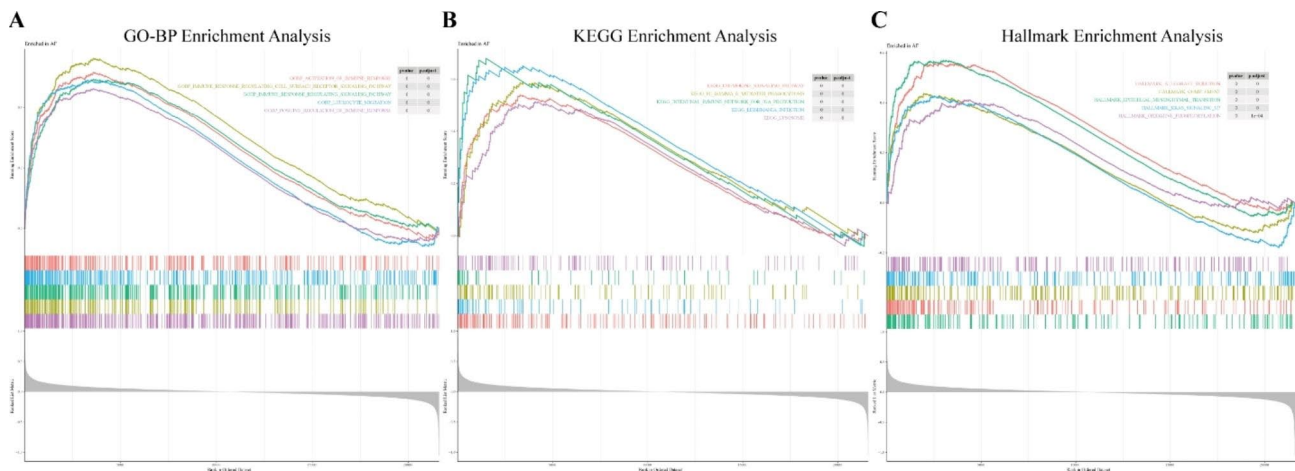


Fig. 5 Functional analysis by GSEA. **(A)** Top 5 GO functions. **(B)** Top 5 KEGG pathways. **(C)** Top 5 Hallmark terms

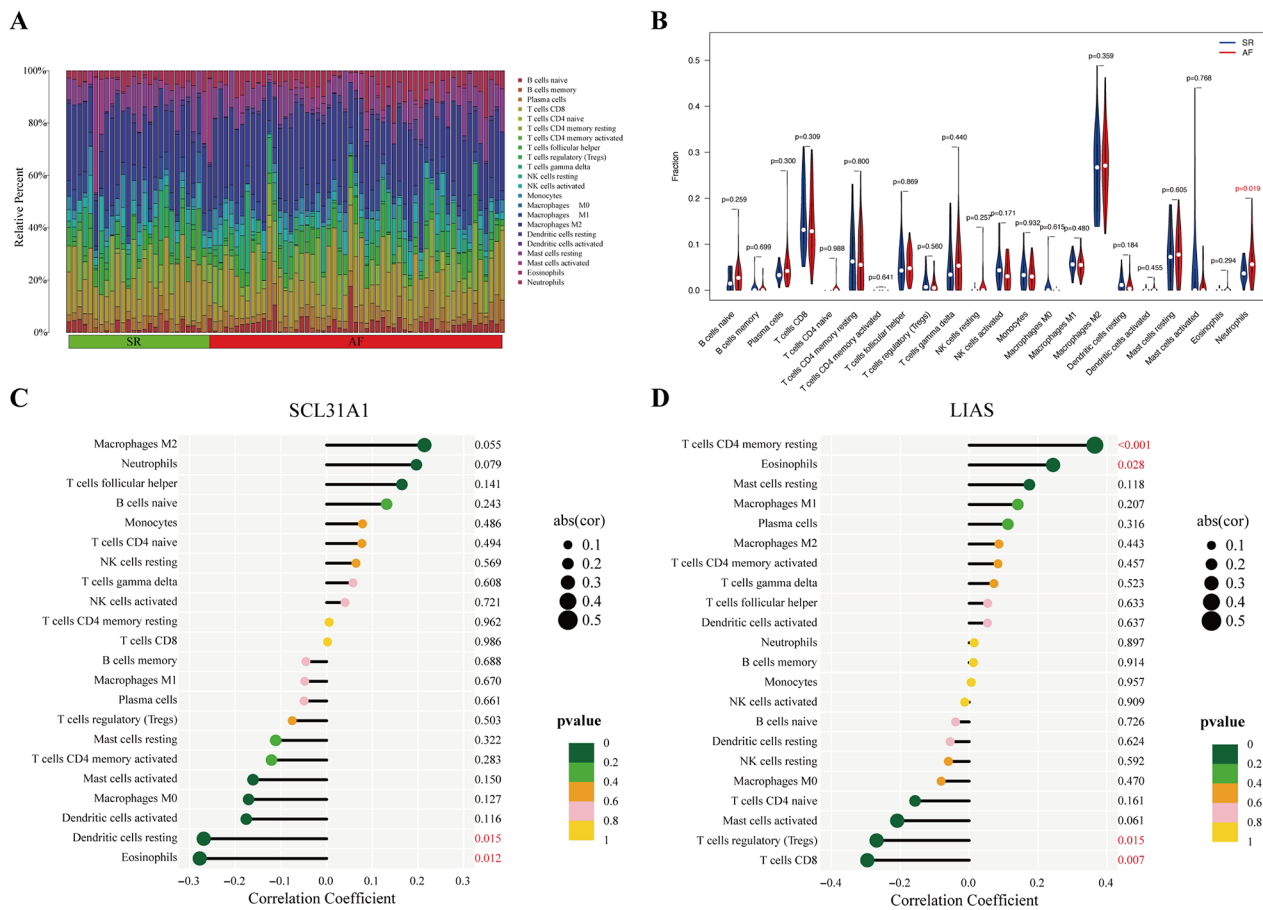


Fig. 6 Immunocyte infiltration and correlation analysis. **(A)** Bar plot displaying the composition of 22 forms of immunocytes between AF and SR samples displayed by different colors. **(B)** Grouped violin plot comparing 22 types of immunocytes between AF patients and SR controls. **(C)** Correlation between SLC31A1 and 22 types of immunocytes. **(D)** Correlation between LIAS and 22 types of immunocytes

causes cuproptosis, a recently recognized type of cell death. [17]. Although previous studies found that not only Cu itself but also WD, a Cu toxicity disease, were related to AF [18, 41], there are still no studies available to suggest any explainable pathogenesis in detail of Cu triggering or maintaining AF. Thus, the purpose of this investigation is not only to identify a novel diagnostic gene signature that may be available to assist clinical diagnosis of AF but also to investigate the relationship between Cu and AF from the aspect of cuproptosis.

This investigation is the first to identify the diagnostic gene signature connected to cuproptosis through bioinformatics methods integrating with EL algorithms, such as RE, XGBoost, and LightGBM algorithms. A cuproptosis-related diagnostic gene signature featuring two genes (SLC31A1 and LIAS) was finally established and validated with good efficacy in identifying AF, specifically with the value of ROC-AUCs and PR-AUCs exceeding 0.75. In contrast, many medical investigations are further required to verify the diagnosis significance of cuproptosis-related diagnostic gene signature. SLC31A1,

also known as “copper transporter 1 (CRT1)”, encodes the protein serving as an increased-affinity Cu importer in the cell membrane. Kim et al. stated that the cardiac-specific knockout of SLC31A1 resulted in morphological, histological, molecular, and physiological hallmarks of cardiomyopathy [42], indicating that SLC31A1 is responsible for preserving the typical cardiac structure and function. LIAS encodes an iron-sulfur enzyme located in the mitochondrion, catalyzing the biosynthesis of lipoic acid. Previous studies reported that alteration of LIAS gene expression affected the development of atherosclerosis [43–45], which is a chronic inflammatory disease and one of the risk factors for AF [46]. Therefore, SLC31A1 and LIAS are likely to impact the AF pathogenesis. Since the direct function of the two genes in AF has been little explored, further studies may focus on the underlying mechanism linking the two cuproptosis-related genes and AF.

In the miRNA-TF-mRNA regulatory network, HOXA9 and TET1 could simultaneously regulate SLC31A1 mRNA and LIAS mRNA. A new investigation by Cai

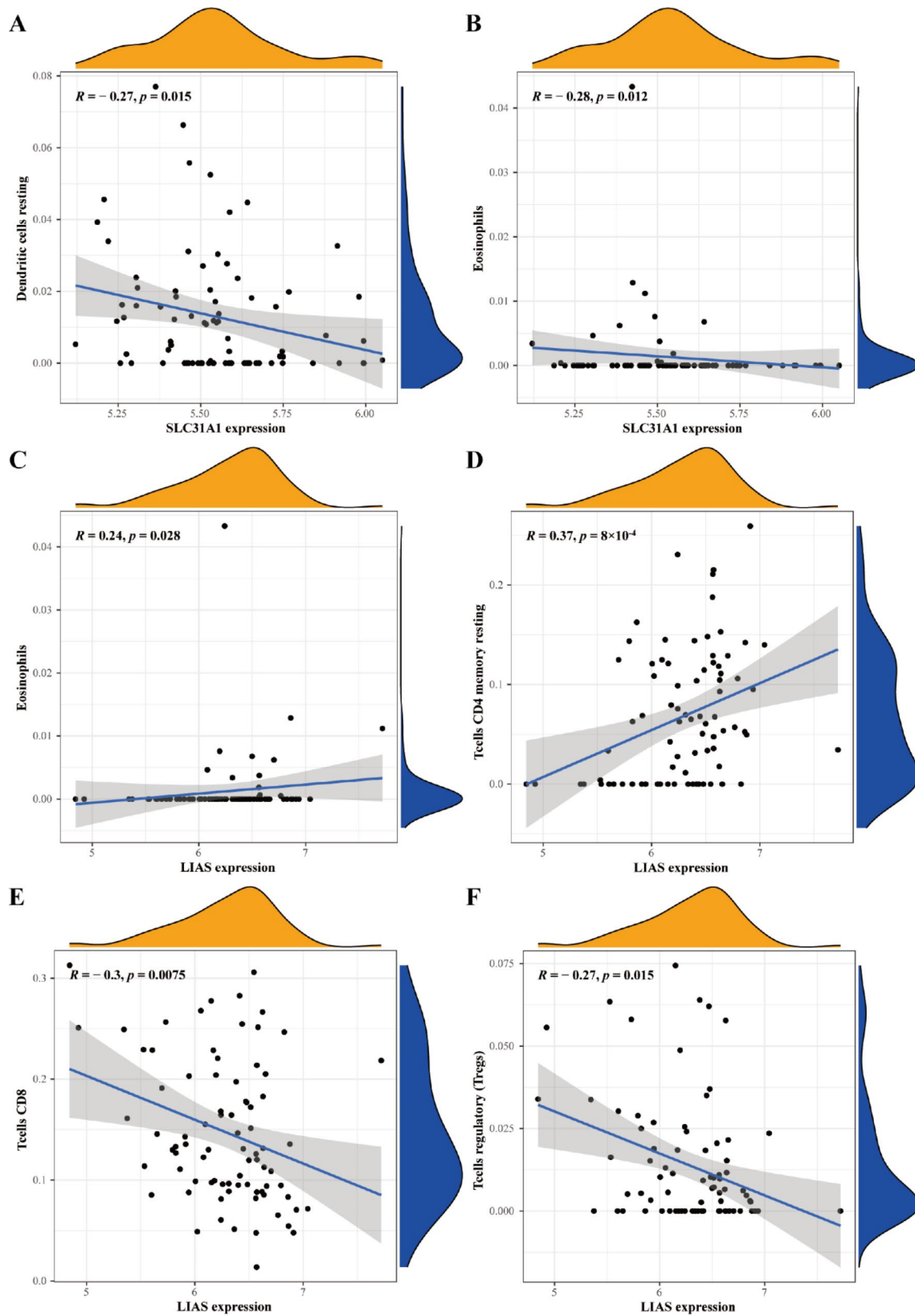


Fig. 7 Correlation analysis of the two genes associated with Cuproptosis and their corresponding infiltrating immunocytes. (A–B) SCL31A1. (C–F) LIAS

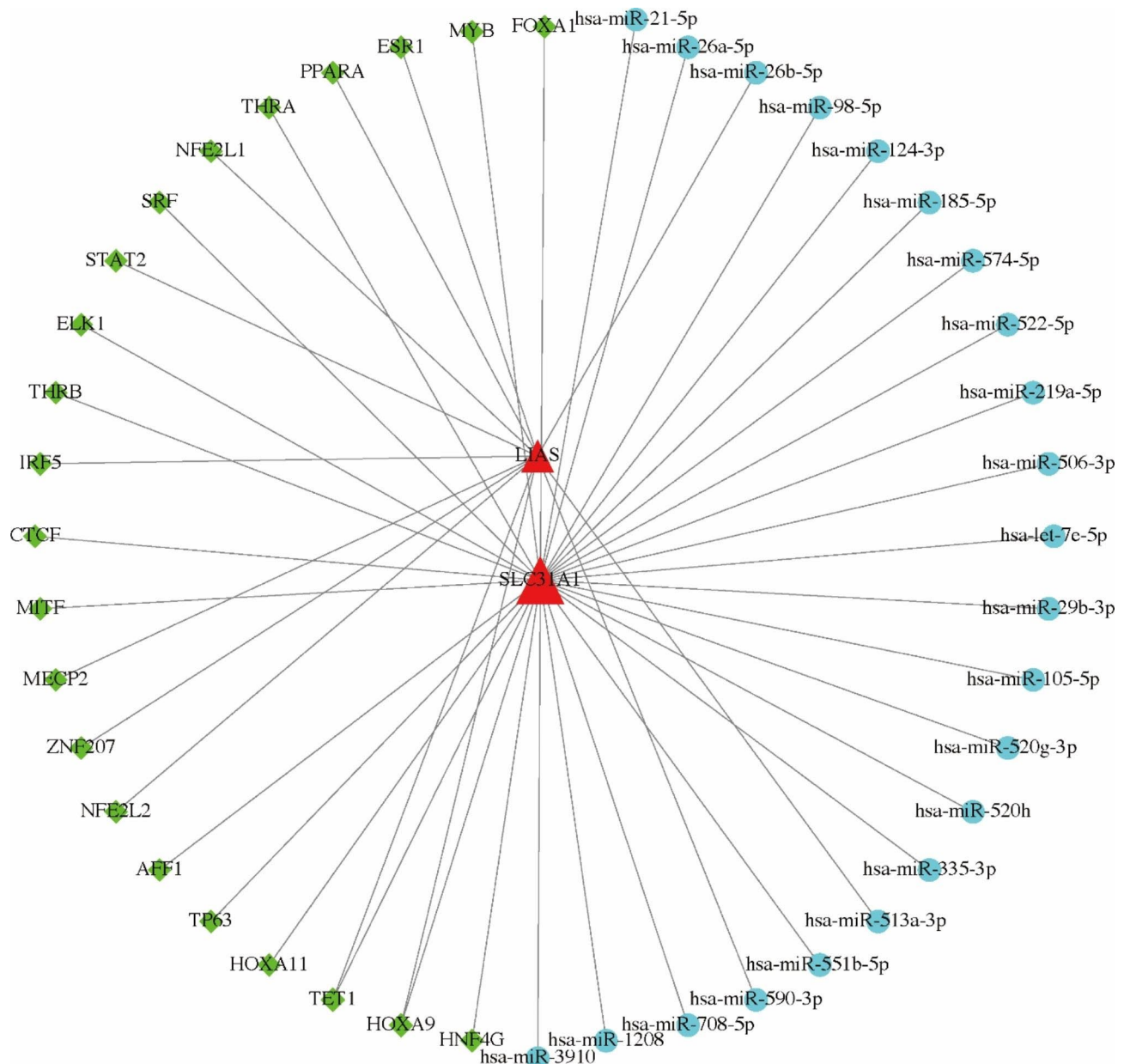


Fig. 8 The miRNA-TF-mRNA network of cuproptosis-related genes. Blue circles represent miRNAs, green diamonds represent TFs, and red triangles represent mRNAs. The edges represent the relationship of miRNA-mRNA or TF-mRNA. The greater the degree of the node, the larger the node

et al. revealed that HOXA9, a member of the Homeobox gene family encoding several greatly conservative progressive transcription factors, could promote cardiomyocyte hypertrophy [47], which is one of the most important structural remodeling features in AF [48]. TET1 is a 5-methylcytosine hydroxylase that initiates the DNA demethylation process [49]. Zhou et al. informed that TET1 was related to the direct cardiac reprogramming of fibroblasts into cardiomyocytes in humans [50]. Since atrial fibrosis involving an abnormal proliferation of cardiac fibroblasts and loss of cardiomyocytes is a characteristic of structural remodeling in AF [51], TET1

can serve as a crucial regulatory point to attenuate structural remodeling by direct cardiac reprogramming in AF. In view of the potential association between the two TFs and AF, it seems reasonable to assume that HOXA9 and TET1 might target SLC31A1 mRNA and LIAS mRNA in the pathogenesis of AF by regulating cuproptosis. However, there is still no study focusing on the interactions between the two TFs and the two cuproptosis-related genes, so further studies will be needed.

The phenomenon of cuproptosis has not been extensively investigated in scientific research. The results of GSEA indicated that various immunocyte-related

functions and mechanisms were significantly enriched in the AF group, which could be specifically identified through the cuproptosis-related gene signature with good performance. Therefore, it is a justifiable hypothesis that cuproptosis may influence the structure of immunocytes infiltrating the atria.

The proportion of neutrophils among individuals with AF was found to be more elevated in comparison to the SR control group. Neutrophils are the most abundant type of leukocyte and have been linked to the regulation of cardiovascular inflammation [52]. The neutrophil-to-lymphocyte ratio elevation is related to the incidence and recurrence of AF [53]. Furthermore, studies showed that neutrophils infiltrating the myocardial interstitium release myeloperoxidase and reactive oxygen species, which induce atrial fibrosis and fibrillation [54]. Furthermore, Babu et al. revealed that the function of neutrophils was sensitive to Cu status [55]. The present findings coincide with previous studies, not only confirming the accuracy of findings but also suggesting the complexity between cuproptosis and neutrophils in AF.

The study conducted a correlation analysis between two genes associated with cuproptosis and infiltrating immunocytes. The findings demonstrated that SLC31A1 exhibited a negative relationship with eosinophils and resting dendritic cells. The study found a positive correlation between LIAS and eosinophils in addition to resting memory CD4⁺ T cells, while a negative relationship was detected between LIAS and CD8⁺ T cells and Tregs. A previous study by Tian et al. indicated that LIAS overexpression led to a reduction in CD4⁺ T cell infiltration and an increase in Treg number in peripheral blood in atherosclerosis [45], but similar investigations are not carried out in the case of AF. Given the lack of research, the sophisticated interactions between cuproptosis-related genes and immunocytes should be investigated in depth on the basis of the assumption mentioned previously.

It is crucial to take into account the restrictions of this investigation while interpreting the outcomes. First, the findings were derived exclusively from public databases using bioinformatics methods. Even though the present results were validated with a validation set, further clinical investigations with large sample sizes are essential for fully evaluating the feasibility of results. Second, this investigation represents the initial attempt to elucidate the correlation between cuproptosis and AF. Currently, research focusing on cuproptosis is still scarce. More in vivo or in vitro functional experiments are required to explore the underlying pathways that link cuproptosis to AF based on the results of this investigation. Third, a bioinformatics approach was used in this study to examine miRNA-TF-mRNA triple interactions associated with cuproptosis-related AF development. Nevertheless, the

obtained miRNA-TF-mRNA interactions still require further experimental verification.

Conclusion

In summary, this investigation demonstrated that cuproptosis was closely related to AF. A 2-gene diagnostic signature that includes cuproptosis-related genes (SLC31A1 and LIAS) based on LightGBM was constructed, and its good performance in the specific recognition of AF was validated. Cuproptosis may be regulated by HOXA9 and TET1 in AF. Moreover, cuproptosis and immunity may orchestrate the pathogenesis of AF. This comprehensive analysis provides the possibility to improve the diagnosis for patients with AF and provides a theoretical base for upcoming research on the associations between immunity and cuproptosis-related genes in AF.

Abbreviations

AF	Atrial fibrillation
ATP7A	ATPase copper transporting Alpha
AUC	Area under the curve
BP	Biological progress
DLST	Dihydrolipoamide s-succinyltransferase
EL	Ensemble learning
GCSH	Glycine cleavage system protein H
GEO	Gene Expression Omnibus
GO	Gene ontology
GSEA	Gene set enrichment analysis
HOXA9	Homeobox A9
KEGG	Kyoto Encyclopedia of Genes and Genomes
LIAS	Lipoic acid synthetase
LightGBM	Light Gradient Boosting Machine
ML	Machine learning
OAC	Oral anticoagulation
PCA	Principal component analysis
PR	precision-recall
RF	Random Forests
ROC	Receiver operator characteristic
SLC31A1	Solute carrier family 31 member 1
SR	Sinus rhythm
TET1	Tet methylcytosine dioxygenase 1
TF	Transcription factor
WD	Wilson's disease
XGBoost	eXtreme Gradient Boosting

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s41065-023-00297-6>.

Supplementary Material 1

Acknowledgements

The authors acknowledge GEO for providing the open-access platform for a large amount of data, as well as their contributors for uploading the meaningful datasets of AF.

Authors' contributions

Qiangsun Zheng and Xinghua Qin contributed to the conception and supervision of this study and provided funding for this research. Yixin Wang designed the study. Yixin Wang, Qiaozhu Wang, Peng Liu and Lingyan Jin performed the data acquisition, bioinformatics analysis, and interpretation of the results. Yixin Wang wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (Grant number 82270335 and 31871172) and the Key Research and Development Project of Shaanxi Province (Grant number 2021SF-132 and 2023-YBSF-600).

Availability of supporting data

The datasets supporting the findings of this study are open-access from GEO (<https://www.ncbi.nlm.nih.gov/geo/>).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 22 March 2023 / Accepted: 2 August 2023

Published online: 24 August 2023

References

- Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, et al. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic surgery (EACTS): the Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. *Eur Heart J*. 2021;42(5):373–498.
- Wolf PA, Dawber TR, Thomas HE Jr, Kannel WB. Epidemiologic assessment of chronic atrial fibrillation and risk of stroke: the Framingham study. *Neurology*. 1978;28(10):973–7.
- Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke: the Framingham Study. *Stroke*. 1991;22(8):983–8.
- Friberg L, Rosenqvist M, Lindgren A, Terént A, Norrving B, Asplund K. High prevalence of atrial fibrillation among patients with ischemic stroke. *Stroke*. 2014;45(9):2599–605.
- Marini C, De Santis F, Sacco S, Russo T, Olivieri L, Totaro R, et al. Contribution of atrial fibrillation to incidence and outcome of ischemic stroke: results from a population-based study. *Stroke*. 2005;36(6):1115–9.
- Hart RG, Pearce LA, Aguilar MI. Meta-analysis: antithrombotic therapy to prevent stroke in patients who have nonvalvular atrial fibrillation. *Ann Intern Med*. 2007;146(12):857–67.
- Kleindorfer DO, Towfighi A, Chaturvedi S, Cockroft KM, Gutierrez J, Lombardi-Hill D, et al. 2021 Guideline for the Prevention of Stroke in patients with stroke and transient ischemic attack: a Guideline from the American Heart Association/American Stroke Association. *Stroke*. 2021;52(7):e364–e467.
- Sanna T, Diener HC, Passman RS, Di Lazzaro V, Bernstein RA, Morillo CA, et al. Cryptogenic stroke and underlying atrial fibrillation. *N Engl J Med*. 2014;370(26):2478–86.
- Jaakkola J, Mustonen P, Kiviniemi T, Hartikainen JEK, Palomäki A, Hartikainen P, et al. Stroke as the First Manifestation of Atrial Fibrillation. *PLoS ONE*. 2016;11(12):e0168010.
- Roselli C, Rienstra M, Ellinor PT. Genetics of Atrial Fibrillation in 2020: GWAS, genome sequencing, polygenic risk, and Beyond. *Circ Res*. 2020;127(1):21–33.
- Zou R, Zhang D, Lv L, Shi W, Song Z, Yi B, et al. Bioinformatic gene analysis for potential biomarkers and therapeutic targets of atrial fibrillation-related stroke. *J Transl Med*. 2019;17(1):45.
- Wang X, Li H, Zhang A, Zhang Y, Li Z, Wang X, et al. Diversity among differentially expressed genes in atrial appendages of atrial fibrillation: the role and mechanism of SPP1 in atrial fibrosis. *Int J Biochem Cell Biol*. 2021;141:106074.
- Aimé-Sempé C, Folliguet T, Rucker-Martin C, Krajewska M, Krajewska S, Heimbürger M, et al. Myocardial cell death in fibrillating and dilated human right atria. *J Am Coll Cardiol*. 1999;34(5):1577–86.
- Yuan Y, Zhao J, Gong Y, Wang D, Wang X, Yun F, et al. Autophagy exacerbates electrical remodeling in atrial fibrillation by ubiquitin-dependent degradation of L-type calcium channel. *Cell Death Dis*. 2018;9(9):873.
- Fu Y, Jiang T, Sun H, Li T, Gao F, Fan B, et al. Necroptosis is required for atrial fibrillation and involved in aerobic exercise-conferred cardioprotection. *J Cell Mol Med*. 2021;25(17):8363–75.
- Dai C, Kong B, Qin T, Xiao Z, Fang J, Gong Y, et al. Inhibition of ferroptosis reduces susceptibility to frequent excessive alcohol consumption-induced atrial fibrillation. *Toxicology*. 2022;465:153055.
- Tsvetkov P, Coy S, Petrova B, Dreishpoon M, Verma A, Abdusamad M, et al. Copper induces cell death by targeting lipoylated TCA cycle proteins. *Science*. 2022;375(6586):1254–61.
- Grandis DJ, Nah G, Whitman IR, Vittinghoff E, Dewland TA, Olgin JE, et al. Wilson's Disease and Cardiac Myopathy. *Am J Cardiol*. 2017;120(11):2056–60.
- Dzieżyc-Jaworska K, Litwin T, Członkowska A. Clinical manifestations of Wilson disease in organs other than the liver and brain. *Ann Transl Med*. 2019;7(Suppl 2):62.
- Percival SS. Copper and immunity. *Am J Clin Nutr*. 1998;67(5 Suppl):1064s–8s.
- Zhou X, Dudley SC. Jr. Evidence for inflammation as a driver of Atrial Fibrillation. *Front Cardiovasc Med*. 2020;7:62.
- Liu Y, Shi Q, Ma Y, Liu Q. The role of immune cells in atrial fibrillation. *J Mol Cell Cardiol*. 2018;123:198–208.
- de la Sánchez AM, Atienza F, Bermejo J, Fernández-Avilés F. Artificial intelligence for a personalized diagnosis and treatment of atrial fibrillation. *Am J Physiol Heart Circ Physiol*. 2021;320(4):H1337–47.
- Yang P, Hwa Yang Y, B Zhou B, Zomaya Y. A review of ensemble methods in bioinformatics. *Curr Bioinform*. 2010;5(4):296–308.
- Breiman L. Random forests. *Mach Learn*. 2001;45(1):5–32.
- Chen T, Guestrin C. Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining 2016. p. 785 – 94.
- Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W et al. Lightgbm: a highly efficient gradient boosting decision tree. *Adv Neural Inf Process Syst*. 2017;30.
- Liaw A, Wiener M. Classification and regression by randomForest. *R news*. 2002;2(3):18–22.
- Zhang Y, Zhang R, Ma Q, Wang Y, Wang Q, Huang Z, et al. A feature selection and multi-model fusion-based approach of predicting air quality. *ISA Trans*. 2020;100:210–20.
- Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 2012;28(6):882–3.
- Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007;8(1):118–27.
- Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12(1):77.
- Wickham H. ggplot2: elegant graphics for data analysis. springer; 2016.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545–50.
- Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omic*. 2012;16(5):284–7.
- Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453–7.
- Tashiro K, Horita N, Nagai K, Ikeda M, Shinkai M, Yamamoto M, et al. HbA1c level cannot predict the treatment outcome of smear-positive non-multi-drug-resistant HIV-negative pulmonary tuberculosis inpatients. *Sci Rep*. 2017;7:46488.
- Ru Y, Kechris KJ, Tabakoff B, Hoffman P, Radcliffe RA, Bowler R, et al. The multiMiR R package and database: integration of microRNA–target interactions along with their disease and drug associations. *Nucleic Acids Res*. 2014;42(17):e133.
- Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. 2013;14:128.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498–504.
- Yan Y-Q, Zou L-J. Relation between zinc, copper, and Magnesium concentrations following cardiopulmonary bypass and postoperative atrial fibrillation

- in patients undergoing coronary artery bypass grafting. *Biol Trace Elem Res.* 2012;148(2):148–53.
42. Kim B-E, Turski ML, Nose Y, Casad M, Rockman HA, Thiele DJ. Cardiac Copper Deficiency activates a systemic signaling mechanism that communicates with the Copper Acquisition and Storage Organs. *Cell Metabol.* 2010;11(5):353–63.
 43. Yi X, Xu L, Kim K, Kim HS, Maeda N. Genetic reduction of lipoic acid synthase expression modestly increases atherosclerosis in male, but not in female, apolipoprotein E-deficient mice. *Atherosclerosis.* 2010;211(2):424–30.
 44. Yi X, Xu L, Hiller S, Kim HS, Maeda N. Reduced alpha-lipoic acid synthase gene expression exacerbates atherosclerosis in diabetic apolipoprotein E-deficient mice. *Atherosclerosis.* 2012;223(1):137–43.
 45. Tian S, Nakamura J, Hiller S, Simington S, Holley DW, Mota R et al. New insights into immunomodulation via overexpressing lipoic acid synthase as a therapeutic potential to reduce atherosclerosis. *Vascul Pharmacol.* 2020;133–4:106777.
 46. da Silva RMFL. Influence of inflammation and atherosclerosis in Atrial Fibrillation. *Curr Atheroscler Rep.* 2017;19(1):2.
 47. Cai S, Liu R, Wang P, Li J, Xie T, Wang M, et al. PRMT5 prevents cardiomyocyte hypertrophy via symmetric dimethylating HoxA9 and repressing HoxA9 expression. *Front Pharmacol.* 2020;11:600627.
 48. Allesie M, Ausma J, Schotten U. Electrical, contractile and structural remodeling during atrial fibrillation. *Cardiovasc Res.* 2002;54(2):230–46.
 49. Liu W, Wu G, Xiong F, Chen Y. Advances in the DNA methylation hydroxylase TET1. *Biomark Res.* 2021;9(1):76.
 50. Zhou Y, Liu Z, Welch JD, Gao X, Wang L, Garbutt T, et al. Single-cell transcriptomic analyses of cell fate transitions during human cardiac reprogramming. *Cell Stem Cell.* 2019;25(1):149–64e9.
 51. Dzeshka MS, Lip GY, Snezhitskiy V, Shantsila E. Cardiac fibrosis in patients with Atrial Fibrillation: mechanisms and clinical implications. *J Am Coll Cardiol.* 2015;66(8):943–59.
 52. Silvestre-Roig C, Braster Q, Ortega-Gomez A, Soehnlein O. Neutrophils as regulators of cardiovascular inflammation. *Nat Rev Cardiol.* 2020;17(6):327–40.
 53. Weymann A, Ali-Hasan-Al-Saegh S, Sabashnikov A, Popov AF, Mirhosseini SJ, Liu T, et al. Prediction of New-Onset and Recurrent Atrial Fibrillation by Complete Blood Count tests: a comprehensive systematic review with Meta-analysis. *Med Sci Monit Basic Res.* 2017;23:179–222.
 54. Friedrichs K, Baldus S, Klinke A. Fibrosis in Atrial Fibrillation - Role of reactive species and MPO. *Front Physiol.* 2012;3:214.
 55. Babu U, Failla ML. Copper status and function of neutrophils are reversibly depressed in marginally and severely copper-deficient rats. *J Nutr.* 1990;120(12):1700–9.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.